



IN THE UNITED STATES PATENT AND TRADE MARK OFFICE

VERIFICATION OF TRANSLATION

I, Anne Desaix

do
hereby declare that I am conversant with the English and French
languages and that I am a competent translator thereof ;

I verify that the attached English translation is a true and correct
translation of the attached French Patent Application No. 9013579 filed
on October 31, 1990.

I further declare that all statements made herein of my own knowledge
are true and that all statements made on information and belief are
believed to be true; and further that these statements were made with
the knowledge that willful false statements and the like so made are
punishable by fine or imprisonment or both under Section 1001 of Title
18 of the United States code and that such willful false statements may
jeopardize the validity of this application or any patent issued thereon.

Date: November 7, 2000

VERIFICATION OF A TRANSLATION
(VERIFICATION D'UNE TRADUCTION)



RECEIVED

JAN 23 2001

TECH CENTER 1600/25

I, the below named translator, hereby declare that :

My name and post office address are as stated below :

That I am knowledgeable in the English language and in the language in which the below identified french Application was filed, and that I believe the English translation of the French Application N° 9013579 filed on October 31, 1990, in the name of INSTITUT PASTEUR

is a true and complete translation of the above identified French Application as filed.

Date: 28/09/1994

Full name of the translator
(nom et prénom du traducteur)

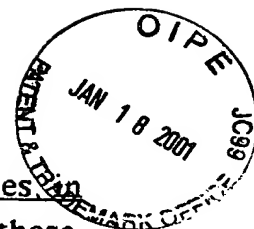
JARVIS, DEREK

Signature of the translator
(Signature du Traducteur)

Derek Jarvis

Post Office Address
(adresse postale)

7, rue d'UPAL
67000 STRASBOURG



Polypeptides implicated in the expression of resistance to glycopeptides, in particular in Gram-positive bacteria. Nucleotide sequence coding for these polypeptides and use for diagnosis

RECEIVED

JAN 23 2001

TECH CENTER 1600/2800

The invention relates to the polypeptides associated with the expression of resistance to antibiotics of the glycopeptide family, in particular in the Gram-positive bacteria, in particular in the family of the Gram-positive cocci. The invention also relates to a nucleotide sequence coding for these polypeptides. It also relates to the use of these polypeptides and their nucleotide sequence as agents for the in vitro detection of resistance to glycopeptides. Among the Gram-positive cocci, the invention relates more particularly to the enterococci, the streptococci and the staphylococci which are of particular importance for the implementation of the invention.

The glycopeptides, which include vancomycin and teicoplanin, are antibiotics which inhibit the synthesis of the bacterial cell wall. These antibiotics are very much used for the treatment of severe infections due to Gram-positive cocci (enterococci, streptococci and staphylococci) in particular in cases of allergy and resistance to the penicillins. In spite of the long clinical usage of vancomycin, this antibiotic has remained active towards almost all of the strains up to 1986, the date at which the first resistant strains were isolated. Since then, resistance to the glycopeptides has been detected by many microbiologists in Europe and in the United States, in particular in strains isolated from immunodepressive patients, making necessary a systematic evaluation of the sensitivity of the microbes in hospital environments.

The activity of the glycopeptides depends on the formation of a complex between the antibiotic and the precursor of the peptidoglycan, more than on the direct interaction with enzymes involved in cell wall metabolism. In particular, it has been observed that the glycopeptides bind to the terminal D-alanyl-D-alanine residues (D-ala-D-ala) of the precursors of the peptidoglycan.

The recent emergence of resistance to the glycopeptides, in particular in the enterococci, has led to certain results being obtained concerning the identification of the factors conferring this resistance.

For example, it has been observed in a particular strain of enterococci, Enterococcus faecium BM4147, that the determinant of resistance to the glycopeptides is localized on a plasmid of 34 kb, the

plasmid pIP816 which has been cloned in E. coli (Brisson Noel et al., 1990, Antimicrob Agents Chemother 34, 924-927).

According to the results obtained hitherto, the resistance to the glycopeptides is associated with the production of a protein of molecular weight of about 40 kDa, the synthesis of this protein being induced by sub-inhibitory concentrations of certain glycopeptides such as vancomycin.

By carrying out a more detailed study of the resistance of certain strains of Gram-positive cocci towards glycopeptides, in particular vancomycin and teicoplanin, the inventors have observed that this resistance would appear to be linked to the expression of several proteins or polypeptides encoded in sequences usually borne by plasmids in the resistant strains. The latest results obtained by the inventors also make it possible to distinguish the genes coding for two phenotypes of resistance, on the one hand, strains highly resistant to the glycopeptides and, on the other, strains with a low level of resistance.

By strain with a high level of resistance is meant a strain of bacteria, in particular a strain of Gram-positive cocci, for which the minimal inhibitory concentrations (MIC) of vancomycin and teicoplanin are higher than 32 and 8 µg/ml, respectively. The MIC of vancomycin towards strains with low-level resistance are included between 16 and 32 µg/ml. These strains are apparently sensitive to teicoplanin.

The inventors have isolated and purified, among the components necessary for the expression of the resistance to the glycopeptides, a particular protein designated VANA which exhibits a certain homology with D-alanine-D-alanine ligases. VANA is nonetheless functionally distinct from the ligases.

The invention relates to polypeptides or proteins implicated in the expression of resistance to antibiotics of the glycopeptide family and, in particular to vancomycin and/or teicoplanin, as well as to the nucleotide sequences coding for such complexes.

The invention also relates to nucleotide probes which can be used for the detection of resistance to the glycopeptides, in particular by means of the polymerase chain reaction (PCR), or by assays involving antibodies.

The polypeptides implicated in the expression of the resistance to the glycopeptides according to the invention are characterized in that they comprise at least 3 proteins or any part of one or more of these proteins necessary to confer on Gram-positive bacteria resistance to antibiotics of the glycopeptide family, in particular to vancomycin and/or teicoplanin, or

to promote this resistance, in particular in strains of the family of the Gram-positive cocci, these proteins or parts of proteins being recognized by antibodies directed against one of the sequences identified in the list of sequences by SEQ ID NO 1, SEQ ID NO 2, SEQ ID NO 3.

These sequences are also designated as ORF1, VANA (or ORF2), ORF3; they characterize the resistance proteins such as are produced by the strain Enterococcus faecium BM4147 described by Leclercq et al. (N. Engl. J. Med. 319 : 157-161).

By the expression "polypeptides" is meant any sequence of amino acids constituting proteins or being of a size smaller than that of a protein.

The expression of resistance to glycopeptides can be expressed by the persistence of an infection due to microbes usually sensitive to the glycopeptides.

A polypeptide or a protein is necessary for the expression of resistance to the glycopeptides, if its absence makes the strain which contains it more sensitive to the glycopeptides and provided that this polypeptide is not present in the sensitive strains.

Different levels of resistance to the glycopeptides exist among the strains of Gram-positive cocci, in particular.

According to a preferred embodiment of the invention, the polypeptides included in the above definitions correspond to the combination of the proteins identified in the list of the sequences by SEQ ID NO 1, SEQ ID NO 2, SEQ ID NO 3.

The inventors have thus observed that the expression of resistance to the glycopeptides in the Gram-positive bacteria requires the expression of at least three proteins or of polypeptides derived from these proteins.

According to a first particular embodiment of the invention, the polypeptides are also characterized in that the amino acid sequences necessary for the expression of resistance to antibiotics of the glycopeptide family are under the control of regulatory elements, in particular proteins corresponding to the sequences designated by SEQ ID NO 4 or SEQ ID NO 5 in the list of the sequences, and which correspond to a regulatory sequence R and a sensor sequence S, respectively.

These regulatory sequences are capable in particular of increasing the level of resistance, to the extent to which they promote the expression of the proteins responsible for resistance comprised in the polypeptides of the invention.

According to another advantageous embodiment of the invention, the above polypeptides are encoded in the sequence SEQ ID NO 6 identified in the list of the sequences which represent the sequences coding for the 5 proteins previously described.

The invention also relates to the purified proteins belonging to the polypeptides previously described. In particular, the invention relates to the purified protein VANA, characterized in that it corresponds to the amino acid sequence SEQ ID NO 2 in the list of the sequences.

The VANA protein contains 343 amino acids and has a calculated molecular weight of 37400 Da.

Other proteins of interest in the framework of the invention correspond to the sequences identified by SEQ ID NO 1, SEQ ID NO 2, SEQ ID NO 4, SEQ ID NO 5 in the list of the sequences.

The invention also relates to any combination of these different proteins in a resistance complex, as well as hybrid proteins composed of one or more of the above proteins, in combination with a defined sequence of amino acids.

Also included in the framework of the invention are the nucleotide sequences coding for one of the amino acid sequences described above.

A particular sequence is the nucleotide sequence of about 7.3 kb, corresponding to the HindIII-EcoRI restriction fragment, such as that obtained starting from the plasmid pIP816 described in the publication of Leclercq et al. - 1988, cited above.

This sequence of 7.3 kb consists of the nucleotide sequence coding for the 3 resistance proteins and the 2 regulatory proteins referred to above. This coding sequence is included in an internal BglIII-XbaI fragment.

The invention also relates to any nucleotide fragment containing the above-mentioned restriction fragment as well as any part of the HindIII-EcoRI, in particular the EcoRI-XbaI fragment of about 3.4 kb coding for the 3 resistance proteins or the EcoRV-SacII fragment of about 1.7 kb coding for VANA or also the HindIII-EcoRI fragment of about 3.3 kb coding for the 2 regulatory proteins.

Another definition of a nucleotide sequence of the invention corresponds to a nucleotide fragment containing the following restriction sites in the following order, such as that obtained from pIP816 mentioned above :

HindIII, BglIII, BglIII, EcoRI, BamHI, XbaI, EcoRI.

Another nucleotide sequence according to the invention is characterized in that it corresponds to the sequence identified by SEQ ID NO 7, or in that it contains this sequence or any part of this sequence capable :

- either of constituting a hybridization probe for the detection of resistance to antibiotics of the glycopeptide family, in particular to vancomycin and/or teicoplanin, especially in strains of the family of the Gram-positive cocci,
- or of coding for a sequence necessary for the expression of resistance to antibiotics of the glycopeptide family, in particular to vancomycin and/or teicoplanin, especially in strains of the family of the Gram-positive cocci.

The sequence SEQ ID NO 7 codes for the 3 resistance proteins mentioned above.

Other preferred nucleotide sequences are the sequences SEQ ID NO 8, SEQ ID NO 9, SEQ ID NO 10, or a variant of one of these sequences provided that it codes for a protein having immunological and/or functional properties similar to those of the proteins encoded in the sequences SEQ ID NO 8, SEQ ID NO 9, SEQ ID NO 10.

These sequences code for the 3 resistance proteins.

The nucleotide sequence designated by SEQ ID NO 8 corresponds to a DNA fragment of 1029 bp situated between the ATG codon at position 377 and the TGA codon at position 1406 on the plasmid pAT214 (Fig.6).

The invention also relates to a nucleotide sequence corresponding to the sequence SEQ ID NO 6 corresponding to the sequence coding for the 5 proteins (2 regulatory proteins and 3 resistance proteins) and also comprising the flanking sequences associated with these coding sequences, or containing this sequence.

Also included in the framework of the invention is a sequence modified with respect to SEQ ID NO 6, characterized in that it lacks the flanking sequences. These flanking sequences are the sequences shown in the following pages and defined as follows :

- sequence upstream from the sequence coding for R: between the bases 1 and 1476,
- sequence between the sequence coding for the sensor protein S and ORF1: between the bases 3347 and 3500,
- sequence downstream from the sequence coding for ORF3: between the bases 6168 and 7227.

The sequence designated by SEQ ID NO 6 is also characterized by the fragment comprising the following restriction sites in the following order :

BglIII - EcoRI - BamHI - EcoRI.

The location of the regulatory proteins and the resistance proteins is shown in Figure 3.

Recombinant sequences characterized in that they comprise one of the above nucleotide sequences also form part of the invention.

The invention also relates to a recombinant vector characterized in that it includes one of the above nucleotide sequences at a site inessential for its replication, under the control of regulatory elements likely to be involved in the expression of the resistance to antibiotics of the glycopeptide family, in particular to vancomycin or teicoplanin, in a defined host.

Particularly advantageous recombinant vectors for the implementation of the invention are the following vectors : pAT214 containing the EcoRV-SacII fragment of 1761 bp containing a nucleotide sequence coding for the VANA protein; in these vectors the sequences of the invention are advantageously placed under the control of promoters such as the lac promoter.

The invention also relates to a recombinant cell host containing a nucleotide sequence such as that previously described or a vector such as that described above under conditions which allow the expression of resistance to antibiotics of the glycopeptide family, in particular resistance to vancomycin and/or teicoplanin, this host being for example selected from the bacteria, in particular the Gram-positive cocci.

For some applications it is also possible to use yeasts, fungi, insect or mammalian cells.

The invention also relates to a nucleotide probe characterized in that it is capable of hybridizing with a sequence previously described, this probe being labelled if necessary. These probes may or may not be specific for the proteins showing resistance to the glycopeptides.

Labels which can be used for the requirements of the invention are the known radioactive labels as well as other labels such as enzymatic labels or chemoluminescent labels.

Probes thus labelled may be used in hybridization tests in order to detect resistance to the glycopeptides in Gram-positive bacteria. In this case, conditions of low stringency will be used.

Nucleotide probes according to the invention may be characterized in that they are specific in Gram-positive bacteria for the sequences coding for a resistance protein to the glycopeptides, in particular to vancomycin and/or teicoplanin, these probes, in addition, recognizing all of these sequences.

By these specific probes is meant any oligonucleotide hybridizing with a nucleotide sequence coding for one of the proteins according to the invention, such as that described in the preceding pages, and not exhibiting a cross-hybridization reaction or amplification reaction (PCR) with sequences present in all of the sensitive strains.

The universal character of the oligonucleotides which can be used in PCR is defined by their capacity to promote specifically the amplification of a nucleotide sequence implicated in resistance in any one strain of Gram-positive bacteria, resistant to the antibiotics of the glycopeptide family.

The size of the nucleotide probes according to the invention may vary depending on the intended use. For the oligonucleotides which can be used in PCR recourse will be had to fragments of a length which is usual in this procedure. In order to construct probes, it is possible to take any part of the sequences of the invention, for example probe fragments of 200 nucleotides.

According to a particular embodiment of the invention, a nucleotide probe is selected for its specificity towards a nucleotide sequence coding for a protein necessary for the expression in Gram-positive bacteria of a high level of resistance to antibiotics of the glycopeptide family, in particular to vancomycin and teicoplanin.

As examples, useful probes will be selected from the intragenic part of the van A gene.

Other particular probes of the invention have the specific character of a nucleotide sequence coding for a protein necessary for the expression in Gram-positive bacteria of a low level of resistance to antibiotics of the glycopeptide family, in particular to vancomycin in Gram-positive bacteria.

It should also be mentioned that oligonucleotide probes which might be derived from the sequence of the van A gene coding for the VANA protein may be used indiscriminately to detect high-level or low-level resistance.

In a particularly preferred manner, a probe of the invention is characterized in that it hybridizes with a non-chromosomal nucleotide sequence of a Gram-positive strain resistant to glycopeptides, in particular to vancomycin and/or teicoplanin, in particular in that it hybridizes with a non-chromosomal nucleotide sequence of a strain of Gram-positive cocci, for example a strain of enterococcus and preferably E.faecium 4147.

In order to distinguish strains with a high level of resistance from strains with a low level of resistance it is possible to carry out a hybridization test using conditions of high stringency.

The oligonucleotides of the invention may be obtained from the sequence of the invention by cutting with restriction enzymes, or by chemical synthesis according to the standard methods.

Furthermore, the invention relates to polyclonal or monoclonal antibodies, characterized in that they recognize the polypeptide(s) described above or an amino acid sequence described above.

These antibodies may be obtained according to the standard methods of antibody production. In particular, in the case of the preparation of the monoclonal antibodies recourse will be had to the method of Köhler and Milstein according to which monoclonal antibodies are prepared by cell fusion between myeloma cells and spleen cells of mice previously immunised with a polypeptide according to the invention, in conformity with the standard procedure.

The antibodies of the invention can advantageously be used for the detection of the presence of proteins characteristic of resistance to the glycopeptides, in particular to vancomycin and teicoplanin.

Particularly useful antibodies are the monoclonal or polyclonal antibodies directed against the VANA protein. Such antibodies advantageously make it possible to detect strains of bacteria, in particular Gram-positive cocci exhibiting high-level resistance to antibiotic of the glycopeptide family.

In order to carry out this detection, recourse will be had to antibodies labelled for example with a radioactive substance or other type of label.

Hence, tests for the detection in Gram-positive bacteria of resistance to the glycopeptides, in particular assays making use of the ELISA procedures, are included in the framework of the invention.

A kit for the in vitro diagnosis of the presence of Gram-positive strains resistant to the glycopeptides, in particular to vancomycin and/or teicoplanin, these strains belonging in particular to the Gram-positive

cocci, for example enterococci, for example E. faecium, is characterized in that it contains :

- antibodies corresponding to the above definition, and labelled where necessary,
- a reagent for the detection of an immunological reaction of the antigen-antibody type.

Furthermore, the agents developed by the inventors offer the very useful advantage of being suitable for the development of a rapid and reliable assay or kit for the detection of Gram-positive strains resistant to the glycopeptides by means of the polymerase chain reaction (PCR). Such an assay makes it possible to improve the sensitivity of the existing tests which remain rather unreliable and, in certain cases, may make possible the detection of all of the representatives of the family of genes coding for proteins responsible for resistance to the glycopeptides in Gram-positive bacteria.

The carrying out of an assay by the method entailing amplification of the genes for these proteins is done by the PCR procedure or by the RPCR (RPCR : abbreviation for reverse polymerase chain reaction).

The RPCR procedure makes possible the amplification of the regions of the genes coding for the NH₂ and COOH terminals which it is desired to detect.

Some specific primers enable the genes of the strains with low-level resistance to be amplified. These primers are selected, for example, from the sequence coding for the resistance protein VANA.

As examples, the following sequences can be used as primers or probes for the detection of an amplification involving the PCR or RPCR method.

P1 : GGX GAA GAT GGX TCX TTX CAA GGX
 G C AG C G
 A

P2 : AAT ACX ATX CCX CGX TTT AC
 C T C
 C

X represents one of the bases A, T, C or G and also corresponds to inosine in all cases.

Naturally, the invention relates to the complementary probes of the oligonucleotides previously described and possibly to the RNA probes which correspond to them.

A kit for the in vitro diagnosis of the presence of strains of Gram-positive bacteria resistant to the glycopeptides, in particular resistant to vancomycin and/or teicoplanin, these strains belonging in particular to the Gram-positive cocci, in particular in that they are strains of enterococci, for example E. faecium, is characterized in that it contains :

- a nucleotide probe complying with the above specifications and where necessary,
- oligonucleoside triphosphates in amounts sufficient for the amplification of the desired sequence,
- a hybridization buffer,
- an agent for polymerizing the DNA.

The invention also relates to a procedure for the in vitro detection of the presence of Gram-positive strains resistant to the glycopeptides, in particular to vancomycin and/or teicoplanin, these strains belonging in particular to the family of the Gram-positive cocci, in particular in that they are strains of enterococci, for example E. faecium, characterized in that it comprises :

- a) the placing of a biological sample likely to contain the resistant strains in contact with a primer constituted by a probe described above, or any part of a sequence previously described, capable of hybridizing with a desired nucleotide sequence necessary for the expression of resistance to the glycopeptides, this sequence being used as matrix in the presence of the 4 different nucleoside triphosphates and a polymerization agent under conditions of hybridization such that for each nucleotide sequence which has hybridized with a primer, an elongation product of each primer complementary to the matrix is synthesized,
- b) the separation of the matrix from the elongation product obtained, this latter then also being capable of behaving as a matrix,
- c) the repetition of step a) so as to produce a detectable amount of the desired nucleotide sequence,
- d) the detection of the amplification product of the nucleotide sequences.

The detection of the elongation products of the desired sequence may be carried out by a probe identical with the primers used to carry out the PCR or RPCR procedure, or also by a probe different from these primers, this probe being labelled if necessary.

Details relating to the implementation of the PCR procedures may be obtained from the patent applications EP 0229701 and EP 0200362.

Other advantages and characteristics of the invention will become apparent in the examples which follow and from the Figures.

Legends to the Figures

- Figure 1 : Electrophoresis on SDS-polyacrylamide gel (SDS-PAGE) of the proteins of the membrane fractions. Lines 1 and 4, molecular weight standards; line 2, *E. faecium* BM4147 placed in culture in the absence of vancomycin; line 3, BM4147 placed in culture with 10 ug/ml of vancomycin. The head of the arrow indicates the position of the VANA protein.
- Figure 2 : A : Restriction maps of the inserts in the plasmids pAT213 and pAT214. The vector and the DNA insert are distinguished by light and dark segments, respectively. The open arrow represents the van A gene.
 B : Strategy for the nucleotide sequencing of the insert of 1761 bp in the plasmid pAT214. The arrows indicate the direction and the extent of the sequencing reactions by the dideoxy method. The synthetic oligonucleotide primer (5' ATGCTCCTGTCTCCTTTC 3' OH) is complementary to the sequence between the positions 361 and 378. Only the pertinent restriction sites are given.
- Figure 3 : position of the sequences R, S, ORF1, ORF2, ORF3.
- Figure 4 : representation of SEQ ID NO 6
- Figure 5 : representation of SEQ ID NO 6 and the corresponding protein.
- Figure 6 : sequence of the VANA gene and the corresponding protein

Materials and methods for the identification and characterization of the van A gene

Bacterial strains and plasmids

The origin of the plasmids used is given in the Table below.

<u>Strain or plasmid</u>	<u>Source or reference</u>
<i>Escherichia coli</i>	
JM83	Messing (1979)
AR1062	Rambach and Hogness (1977)
JM103	Hannshan (1983)
ST640	Lugtenberg and van Schijndel van-Dam (1973)

Enterococcus faecium

BM4147	Leclercq et al. (1988)
Plasmid pUC18	Norrande et al. (1983)
pAT213	Brisson-Noël et al. (1990)
pAT214	This work

Preparation of the enterococcal membranes

Enterococcus faecium BM4147 was cultured in 500 ml of BHI broth until the optical density (OD₆₀₀) reached 0.7. Induction was effected with 10 µg/ml of vancomycin (Eli Lilly Indianapolis Ind). The subsequent steps were performed at 4°C. The cells were recovered by centrifugation for 10 minutes at 6000g, washed with a TE buffer (0.01 M TRIS-HCl, 0.002 M EDTA, pH 7.0) and lysed with glass beads (100 µm in diameter) in a Braun apparatus for 2 minutes. The cell debris were separated by centrifugation for 10 minutes at 6000 g. The membranes were collected by centrifugation for 1 hour at 65000g and resuspended in 0.5 ml of TE buffer.

Preparation of the protein extracts

Plasmids were introduced by transformation into the *E. coli* AR1062 strain prepared in the form of bacterial vesicles. The bacterial vesicles were recovered on sucrose gradients and the proteins were labelled with 50 µCi of ³⁵S/-L-methionine (Amersham, Great Britain) according to the method of Rambach and Hogness (1977, P.N.A.S., USA, 74; 5041-5045).

Preparation of the membrane fractions and the cytoplasmic fractions of *E. coli*

E. coli JM83 and strains derived from it were placed in culture in a BHI medium until an optical density (OD₆₀₀) of 0.7 was attained, they were washed and suspended in a TE buffer. The cell suspension was treated by sonication for 20 seconds with pulses of 50 W in a cell fragmentation apparatus in a Branson B7 sonication apparatus and the intact cells were removed by centrifugation for 10 minutes at 6000g. The supernatant was fractionated into membrane and cytoplasmic fractions by means of centrifugation for 1 hour at 100,000g.

Electrophoresis on SDS-polyacrylamide gel (SDS-PAGE)

The proteins of the bacterial fractions were separated by means of SDS-PAGE on linear gradients of polyacrylamide gels (7.5% - 15%) (Laemmli 1970, Nature 227 : 680-685). The electrophoresis was carried out for 1 hour at 200 V, then for 3 hours at 350 V. The gels were stained with Coomassie blue. The proteins of the extracts were separated on 10% polyacrylamide gels and visualized by means of autoradiography.

Purification of the protein band and determination of the N-terminal sequence

The proteins of the membrane fractions of an induced culture of *E. faecium* BM4147 were separated by means of SDS-PAGE. The gel was electrotransferred during 1 hour at 200 mA to a polyvinylidene difluoride membrane (Immobilon Transfer, Millipore) by using a transfer apparatus (Electrophoresis Unit LKB 2117 Multiphor II) in accordance with the instructions of the manufacturer. The transferred proteins were stained with Ponceau Red. The portion of membrane bearing the protein of interest was cut out, centered on a Teflon filter and placed in the cartridge of a sequenator (Applied Biosystems Sequenator model 470A). The protein was sequenced by means of the automated Edman degradation procedure (1967, Eur. J. Biochem. 1; 80-81).

Construction of the plasmids

The plasmid pAT213 (Brisson-Noël et al., 1990, Antimicrob. Agents Chemother., 34; 924-927) consists of a *EcoRI* DNA fragment of 4.0 kb of the enterococcal plasmid pIP816 cloned at the *EcoRI* site of a Gram-

positive-Gram-negative shuttle vector pAT187 (Trieu-Cuot et al., 1987, FEMS Microbiol. Lett. 48; 289-294). In order to construct pAT214, the EcoRV-SacII DNA fragment of 1761 bp of pAT213 was purified, treated with the Klenow fragment of the DNA polymerase I of E. coli and ligated to the DNA of pUC18 which had previously been digested with SmaI and dephosphorylated (Figure 2). The cloning (Maniatis et al., 1982 Cold Spring Harbor Laboratory Press) was carried out with restriction endonucleases (Boehringer Mannheim and Pharmacia), with the T4 DNA ligase (Pharmacia) and alkaline phosphatase (Pharmacia) according to the instructions of the manufacturer.

Subcloning in M13 and nucleotide sequence

The DNA restriction fragments were subcloned in the polylinker of the replicative forms of the derivatives mp18 and mp19 of the bacteriophage M13 (Norranders et al., 1983, Gene 26; 101-106), obtained from Pharmacia P-L Biochemicals. E. coli JM103 was transfected with recombinant phages and the single-stranded DNA was prepared. The nucleotide sequencing was performed by the dideoxy chain termination method (Sanger et al., 1977, P.N.A.S. USA 74; 5463-5467) by using a T7 DNA polymerase (Sequenase, United States Biochemical Corporation, Cleveland Ohio) and γ -³⁵S/dATP (Amersham, Great Britain). The reaction products were revealed on 6% polyacrylamide gels containing a denaturing buffer.

Data-processing analysis and data on the sequence

The complete DNA sequence was assembled by using the computer programs DBCOMP and DBUTIL (Staden, 1980, Nucleic Acids Res. 8; 3673-3694). The protein data bank PSEQIP of the Pasteur Institute was screened using an algorithm developed by Claverie (1984, Nucleic Acids Res. 12; 397-407). The alignments between pairs of amino acid sequences were constructed using the algorithm of Wilbur et al. (1983, P.N.A.S. USA 80; 726-730). The statistical significance of the homology was evaluated with the algorithm of Lipman and Pearson (1985, Science 227; 1435-1440).

For each comparison 20 amino acid sequences were used to calculate the mean values and the standard deviations from the random results

Genetic complementation tests

The plasmids were introduced by transformation into E. coli ST640, a temperature-sensitive mutant with an unmodified D-ala-D-ala ligase (Lugtenberg et al., 1973, J. Bacteriol. 110; 26-34). The transformants were selected at 30°C on plates containing 100 ug/ml of ampicillin and the presence of the plasmid DNA of the expected size and the restriction maps were confirmed. Single colonies grown at 30°C in BHI broth containing ampicillin were placed on a BHI agar medium containing both 100 ug/ml of ampicillin and 50 uM of isopropyl-1-thio-beta-D-galacto-pyranoside (IPTG) and the plates were incubated at a permissive temperature of 30°C and at a non-permissive temperature of 42°C. The complementation test was considered to be positive if the colonies were present after 18 hours of incubation at 42°C.

Results

Identification of the VANA protein and its N-terminal sequence

The membrane fractions of the E. faecium BM4147 cells placed in culture, on the one hand, under conditions of induction and, on the other, in the absence of induction, were analysed by means of SDS-PAGE. The sole detectable difference associated with the exposure to sub-inhibitory concentrations of vancomycin was the marked intensification of a band which corresponded to a protein of an estimated molecular weight of about 40 kDa. In the induced cells and in the non-induced cells the protein band represents the same protein because this band is absent from membranes of a derivative of BM4147 which has lost the pIP816 plasmid. The inducible protein, designated as VANA, was purified after SDS-PAGE and automated Edman degradation was carried out on a 50 pmol. sample. Nine amino acids of the N-terminal sequence of VANA were identified : Met Asn Arg Ile Lys Val Ala Ile Leu.

Subcloning of the van A gene

The insert of 4.0 kb of the plasmid pAT213 bears the determinant for resistance of E. faecium BM4147 to the glycopeptides. Various restriction fragments of this insert were subcloned in pUC18 and the recombinant plasmids specific for VANA in E. coli were identified by SDS-PAGE analysis of the proteins of the cytoplasmic and membrane fractions or of the extracts of the bacterial vesicles. This approach was used since E. coli

is intrinsically resistant to the glycopeptide. The EcoRV-SacII insert of the pAT214 plasmid (Figure 2) codes for a unique polypeptide of 40 kDa which migrates together with VANA, derived from the membrane preparations of E. faecium BM4147.

Nucleotide sequence of the insert inpAT214 and identification of the van A coding sequence.

The nucleotide sequence of the EcoRV-SacII insert of 1761 bp in pAT214 was determined on both strands of the DNA according to the strategy described in Figure 2. The location of the termination codons (TGA, TAA, TAG) in three reading frames on each DNA strand showed the presence of a unique open reading frame (ORF) which was sufficiently long to code for the VANA protein. This reading frame ORF is located between the TAA codon at position 281 and the TAG codon at position 1406. The amino acid sequence deduced for ORF was compared with that of the N-terminus of VANA. The nine amino acids identified by protein sequencing are encoded in the nucleotide sequence beginning with the ATG (methionine) codon at position 377 (Figure 3). This codon for the initiation of translation is preceded by a sequence (TGAAAGGAGA), characteristic of a ribosomal binding site (RBS) in Gram-positive bacteria which is complementary to 8 bases of the rRNA of the 16S subunit of Bacillus subtilis corresponding to the sequence (3' OH UCUUCCUCC 5') (Moran et al., 1982, Mol. Gen. Genet., 186; 339-346). In this ORF, there is no other ATG or GTG initiation codon between the positions 281 and 377. The sequence of 1029 bp which extends from the ATG codon at position 377 to the TGA codon at position 1406 codes for a protein containing 343 amino acid residues. The calculated molecular weight of this protein is 37400 Da, which is in agreement with the estimate of 40 kDa obtained by SDS-PAGE analysis.

Homology of the amino acid sequences of VANA and the D-ala-D-ala ligase enzymes

The screening of the protein data bank PSEQIP has shown the existence of a sequence homology between VANA and the D-ala-D-ala ligases of E. coli (ECOALA, Robinson et al., 1986, J. Bacteriol. 167; 809_817) and Salmonella typhimurium (DALIG, Daub et al., 1988, biochemistry 27; 3701-3708). the calculated percentage of homology between pairs of proteins was between 28% and 36% for the identical

amino acids and between 48% and 55% on taking into consideration homologous amino acids. VANA and DALIG are more closely related. The statistical significance of these similarities was assessed by aligning VANA and sequences containing the same composition of amino acids as DALIG or ECOALA (Lipman and Pearson, 1985, Science 227; 1435-1440).

Genetic complementation test for D-ala-D-ala ligase activity

The E. coli ST640 strain is a heat-sensitive mutant exhibiting a deficient D-ala-D-ala ligase activity (Lugtenberg et al., 1973, J. Bacteriol. 113 : 96-104). The plasmids pUC18 and pAT214 were introduced into E. coli ST640 by transformation. The strains ST640 and ST640 (pUC18) grew normally only at the permissive temperature (30°C) whereas E. coli ST640 (pAT214) grew at both the permissive temperature and at the non-permissive temperature (42°C).

This test shows that VANA is biologically active in E. coli and is probably capable of catalysing the same ligation reaction as DALIG.

The sequences which form the subject of the invention are given on the following pages after the list of the sequences containing the description of these sequences. In the list of sequences, the proteins are located with respect to the position of the nucleotide bases which correspond to the amino acids of the termini of the proteins.

List of the sequences (contained in the sequence presented below)

- Amino acid sequences

SEQ ID NO 1 : sequence of the first resistance protein , corresponding to the amino acid sequence of the open reading frame No.3, starting at base 3501 and terminating at base 4529.

SEQ ID NO 2 : sequence of the VANA protein , corresponding to the amino acid sequence of the open reading frame No.1, starting at base 4429 and terminating at base 5553.

SEQ ID NO 3 : sequence of the third resistance protein , corresponding to the amino acid sequence of the open reading frame No.3, starting at base 5526 and terminating at base 6167.

SEQ ID NO 4 : sequence of the regulatory protein R, corresponding to the amino acid sequence of the open reading frame No.1, starting at base 1477 and terminating at base 2214.

SEQ ID NO 5 : sequence of the sensor protein S, corresponding to the amino acid sequence of the open reading frame No.2, starting at base 2180 and terminating at base 3346.

- Nucleotide sequences

SEQ ID NO 6 : nucleotide sequence containing the sequence coding for the 5 proteins as well as the flanking sequences

SEQ ID NO 7 : sequence containing the sequence coding for the 3 resistance proteins and the flanking sequences, starting at base 3501 and terminating at base 6167.

SEQ ID NO 8 : sequence of the van A gene, starting at base 4429 and terminating at base 5553.

SEQ ID NO 9 : sequence of the first resistance protein , starting at base 3501 and terminating at base 4529.

SEQ ID NO 10 : sequence of the third resistance protein , starting at base 5526 and terminating at base 6167.

LysLeuPhePheLeuLeuIleCys***ArgPheThrAsnArgIleLys***LeuLeuPhe
 SerPheSerPheCysSerPheValArgAspLeuLeuThrValLeuAsnSerPhePheSer
 AlaPheLeuPheAlaHisLeuLeuGluIleTyr***ProTyr***IleAlaSerPheGln
 AAGCTTTTCTTTTGCTCATTGTTAGAGATTTACTAACCGTATTAAATAGCTTCTTTTC

SerHisCysProCysPheProHisHisSerPheLysCysSerAspSerArgGlnTyrAsn
 AlaIleAlaLeuAlaSerHisThrIleLeuSerSerValValIleAlaGlySerIleIle
 ProLeuProLeuLeuProThrProPhePheGlnVal*****GlnAlaVal***Phe
 AGCCATTGCCCTTGCTTCCCACACCATTTCTTCAAGTGTAGTGATAGCAGGCAGTATAAT

100

PheValPheSer***LysIleTyrAlaPheMetGln***MetAsnGlyIleThrIlePhe
 LeuPhePheLeuArgLysSerMetHisSerCysSerArg***MetAlaSerProPheSer
 CysPhePheLeuGluAsnLeuCysIleHisAlaValAspGluTrpHisHisHisPhePro
 TTTGTTTTTTCTTAGAAAATCTATGCATTCATGCAGTAGATGAATGGCATCACCATTTTC

GlnSer***LeuMetLysValLeuLysCysHisSerIlePheThrGlnGlyLysSerTyr
 LysAlaAsn*****ArgTyrLeuAsnValIleArgTyrSerLeuArgValLysValThr
 LysLeuIleAspGluGlyThr***MetSerPheAspIleHisSerGly***LysLeuGln
 CAAAGCTAATTGATGAAGGTACTTAAATGTCATTTCGATATTCACCTCAGGGTAAAAGTTAC

200

LysValValPheThrSerAsnPhePheGlnMetIleProLysCysIlePheProLeuArg
 LysSerTyrSerLeuArgIleSerPheLys***SerGlnSerValPheSerLeu***Gly
 SerArgIleHisPheGluPheLeuSerAsnAspProLysValTyrPheProPheGluAsp
 AAAGTCGTATTCACTTCGAATTTCTTTCAAATGATCCCAAAGTGTATTTCCCTTTGAGG

300

IleMetIleLysArgGlyTrpThrAsnThrAsnLeuPheArgTyrIleLeuTyrAspArg
 *****SerSerGluAspGlyLeuThrProIleCysPheAspIleTyrCysMetThrGlu
 AsnAspGlnAlaArgMetAsp***HisGlnSerValSerIleTyrIleVal***ProAsn
 ATAATGATCAAGCGAGGATGGACTAACACCAATCTGTTTCGATATATATTGTATGACCGA

 IleTrpAspAlaPheAspMetSerValTrpProThrGlyIleProLysAsnSer***Leu
 SerGlyMetLeuLeuIle***ValTyrGlyGlnProGlyTyrArgArgThrAlaAsn***
 LeuGlyCysPhe***TyrGluCysMetAlaAsnArgAspThrGluGluGlnLeuIleGlu
 ATCTGGGATGCTTTTGATATGAGTGTATGGCCAACCGGGATACCGAAGAACAGCTAATTG
 400
 AsnSerLysSer***ThrValPhePheProProSerLeuIleAsnTyrPhe***IlePro
 ThrAlaAsnProLysArgPheSerSerLeuLeuArgLeuLeuThrIleSerLysSerArg
 GlnGlnIleLeuAsnGlyPheLeuProSerPheAlaTyr***LeuPheLeuAsnProVal
 AACAGCAAATCCTAAACGGTTTTCTTCCCTCCTTCGCTTATTAATACTATTTCTAAATCCCG

 PheGlyLysSerGluValGlyProGlnTyrProPheIlePheArgAspLeuHisLysSer
 LeuGluLysValLys***ValProSerIleHisSerSerSerGlyIleCysIleLysAla
 TrpLysLys***SerArgSerProValSerIleHisLeuGlnGlyPheAla***LysPro
 TTTGGAAAAAGTGAAGTAGGTCCCCAGTATCCATTCATCTTCAGGGATTTCATAAAAGC
 500
 LeuSerLeuPheArgCysLysGlnPheSerThrSerArgAsnPheHisSerValSerPhe
 CysLeuCysSerGlyValSerAsnSerLeuProLeuAlaIlePheIleGlnTyrHisSer
 ValSerValProVal***AlaIleLeuTyrLeuSerGlnPheSerPheSerIleIlePro
 CTGTCTCTGTTCCGGTGTAAGCAATTCTCTACCTCTCGCAATTTTCATTTCAGTATCATTC
 600

HisPheCysIlePheAsnLeuLeuValGlnLeuTyrIleAsnArgValTyrSerIleAsp
 IleSerValPheSerIleTyr***PheAsnTyrIleSerIleGluCysThrLeuLeuIle
 PheLeuTyrPheGlnPheIleSerSerIleIleTyrGln***SerValLeuTyr***Tyr
 CATTTCTGTATTTTCAATTTATTAGTTCAATTATATATCAATAGAGTGTACTCTATTGAT

ThrAsnValValAsp*****AsnHisSer***GluArgLeuIleArgLeuValSerLys
 GlnMet*****ThrAspLysIleIleValLysSerValSer***AspLeuSerGlnLys
 LysCysSerArgLeuIleLysSer***LeuArgAlaSerHisLysThrCysLeuLysAsn
 ACAAATGTAGTAGACTGATAAAATCATAGTTAAGAGCGTCTCATAAGACTTGTCTCAAAA

700

MetArg***TyrPheAlaGluAsnArgLeuTyrSerCysGlnPheAsp***ProGluSer
 ***GlyAspIleLeuArgLysIleGlyTyrIleArgValSerSerThrAsnGlnAsnPro
 GluValIlePheCysGlyLysSerValIlePheValSerValArgLeuThrArgIleLeu
 ATGAGGTGATATTTTGC GGAAAATCGGTTATATTCGTGTCAGTTCGACTAACCAGAATCC

PheLysThrIleSerAlaValGluArgAspArgAsnGlyTyrTyrIleLysArgLysPhe
 SerArgGlnPheGlnGlnLeuAsnGluIleGlyMetAspIleIle***ArgGluSerPhe
 GlnAspAsnPheSerSer***ThrArgSerGluTrpIleLeuTyrLysGluLysValSer
 TTCAAGACAATTTTCAGCAGTTGAACGAGATCGGAATGGATATTATATAAAGAGAAAGTTT

800

GlnGluGlnGlnArgIleAlaSerAsnPheLysLysCys***ThrIleTyrArgLysMet
 ArgSerAsnLysGlySerArgAlaThrSerLysSerValArgArgPheThrGlyArg***
 GlyAlaThrLysAspArgGluGlnLeuGlnLysValLeuAspAspLeuGlnGluAspAsp
 CAGGAGCAACAAAGGATCGCGAGCAACTTCAAAAAGTGTTAGACGATTTACAGGAAGATG

900

ThrSerPheMetLeuGlnThr***LeuGluSerLeuValValHisLysIleTyrLeuAsn
 HisHisLeuCysTyrArgLeuAsnSerAsnHisSer***TyrThrArgSerIle***Ile
 IleIleTyrValThrAspLeuThrArgIleThrArgSerThrGlnAspLeuPheGluLeu
 ACATCATTATGTTACAGACTTAACTCGAATCACTCGTAGTACACAAGATCTATTTGAAT

 SerIleThrTyrGluIleLysArgGlnValAsnHis***LysIleHisGlyLeu
 AsnArg***HisThrArg***LysGlyLysPheLysIleThrLysArgTyrMetAla***
 IleAspAsnIleArgAspLysLysAlaSerLeuLysSerLeuLysAspThrTrpLeuAsp
 TAATCGATAACATACGAGATAAAAAGGCAAGTTTAAAATCACTAAAAGATACATGGCTTG
 1000 . . .
 IleTyrGlnLysIleIleHisThrAlaAsnSer***LeuLeu***TrpLeuValLeuThr
 PheIleArgArg***SerIleGlnProIleLeuAsnTyrCysAsnGlyTrpCys***Pro
 LeuSerGluAspAsnProTyrSerGlnPheLeuIleThrValMetAlaGlyValAsnGln
 ATTTATCAGAAGATAATCCATACAGCCAATTCTTAATTACTGTAATGGCTGGTGTAAACC

 Asn***SerGluIleLeuPheGly***AspAsnValLysGlyLeuAsnTrpLeuArgLys
 IleArgAlaArgSerTyrSerAspGluThrThr***ArgAsp***IleGly***GluArg
 LeuGluArgAspLeuIleArgMetArgGlnArgGluGlyIleGluLeuAlaLysLysGlu
 AATTAGAGCGAGATCTTATTCGGATGAGACAACGTGAAGGGATTGAATTGGCTAAGAAAG
 1100 . . .
 LysGluSerLeuLysValAsp***ArgSerIleIleLysIleThrGlnGlu***IleMet
 ArgLysVal***ArgSerIleLysGluValSer***LysSerArgArgAsnGluLeuCys
 GlyLysPheLysGlyArgLeuLysLysTyrHisLysAsnHisAlaGlyMetAsnTyrAla
 AAGGAAAGTTTAAAGGTCGATTAAAGAAGTATCATAAAAATCACGCAGGAATGAATTATG
 1200 . . .

ArgArgLysLeuTyrLysGluGlyAsnMetThrValAsnGlnIleCysGluIleThrAsn
 GlyGluSerTyrIleLysLysGluIle***Leu***IleLysPheValLysLeuLeuMet
 AlaLysAlaIle***ArgArgLysTyrAspCysLysSerAsnLeu***AsnTyr***Cys
 CGGXXAAAGCTATATAAAGAAGGAAATATGACTGTAAATCAAATTTGTGAAATTACTAAT

ValSerArgAlaSerLeuTyrArgLysLeuSerGluValAsnAsn***ProPheCysIle
 TyrLeuGlyLeuHisTyrThrGlyAsnTyrGlnLys***IleIleSerHisSerValPhe
 Ile***GlyPheIleIleGlnGluIleIleArgSerGlu***LeuAlaIleLeuTyrSer
 GTATCTAGGGCTTCATTATACAGGAAATTATCAGAAGTGAATAATTAGCCATTCTGTATT

1300

ProLeuMetGlyAsnIlePheLysGluGluLysGluThrIleLysTyr***GlnProPro
 Arg***TrpAlaIlePheLeuLysLysLysArgLysLeu***AsnIleAsnSerLeuLeu
 AlaAsnGlyGlnTyrPhe***ArgArgLysGlyAsnTyrLysIleLeuThrAlaSer***
 CCGCTAATGGGCAATATTTTAAAGAAGAAAAGGAACTATAAAATATTAACAGCCTCCT

SerAspAlaGluLysProPheAspLysLysArgIleIleIleLeuArgAsnSer***Ser
 AlaMetProLysSerProLeuIleLysLysGluSerSerSer***GluIleLeuSerHis
 ArgCysArgLysAlaLeu*****LysLysAsnHisHisLeuLysLysPheLeuValIle
 AGCGATGCCGAAAAGCCCTTTGATAAAAAAGAATCATCATCTTAAGAAATTCTTAGTCA

1400

PheIleMet***MetLeuIleAsnSerAlaLeu***SerAspLysLeuLeuArgAlaAsn
 LeuLeuCysLysCysLeu***IleArgProTyrAsnLeuIleAsnTyr***GlyGlnThr
 TyrTyrValAsnAlaTyrLysPheGlyProIleIle*****IleIleLysGlyLysLeu
 TTTATTATGTAAATGCTTATAAATTCGGCCCTATAATCTGATAAATTATTAAGGGCAAAC

1500

LeuCysGluArgValIleThrMetSerAspLysIleLeuIleValAspAspGluHisGlu
 TyrValLysGly*****Leu***AlaIleLysTyrLeuLeuTrpMetMetAsnMetLys
 Met***LysGlyAspAsnTyrGluArg***AsnThrTyrCysGly*****Thr***Asn
 TTATGTGAAAGGGTGATAACTATGAGCGATAAAATACTTATTGTGGATGATGAACATGAA

 IleAlaAspLeuValGluLeuTyrLeuLysAsnGluAsnTyrThrValPheLysTyrTyr
 LeuProIleTrpLeuAsnTyrThr***LysThrArgIleIleArgPheSerAsnThrIle
 CysArgPheGly***IleIleLeuLysLysArgGluLeuTyrGlyPheGlnIleLeuTyr
 ATTGCCGATTTGGTTGAATTATACTTAAAAACGAGAATTATACGGTTTTCAAATACTAT
 1600
 ThrAlaLysGluAlaLeuGluCysIleAspLysSerGluIleAspLeuAlaIleLeuAsp
 ProProLysLysHisTrpAsnVal***ThrSerLeuArgLeuThrLeuProTyrTrpThr
 ArgGlnArgSerIleGlyMetTyrArgGlnVal***Asp***ProCysHisIleGlyHis
 ACCGCCAAAGAAGCATTGGAATGTATAGACAAGTCTGAGATTGACCTTGCCATATTGGAC

 IleMetLeuProGlyThrSerGlyLeuThrIleCysGlnLysIleArgAspLysHisThr
 SerCysPheProAlaGlnAlaAlaLeuLeuSerValLysLys***GlyThrSerThrPro
 HisAlaSerArgHisLysArgProTyrTyrLeuSerLysAsnLysGlyGlnAlaHisLeu
 ATCATGCTTCCCGGCACAAGCGGCCTTACTATCTGTCAAAAAATAAGGGACAAGCACACC
 1700
 TyrProIleIleMetLeuThrGlyLysAspThrGluValAspLysIleThrGlyLeuThr
 IleArgLeuSerCys***ProGlyLysIleGlnArg***IleLysLeuGlnGly***Gln
 SerAspTyrHisAlaAspArgGluArgTyrArgGlyArg***AsnTyrArgValAsnAsn
 TATCCGATTATCATGCTGACCGGGAAAGATACAGAGGTAGATAAAATTACAGGGTTAACA
 1800

IleGlyAlaAspAspTyrIleThrLysProPheArgProLeuGluLeuIleAlaArgVal
 SerAlaArgMetIleIle***ArgSerProPheAlaHisTrpSer***LeuLeuGly***
 ArgArgGly***LeuTyrAsnGluAlaLeuSerProThrGlyValAsnCysSerGlyLys
 ATCGGCGCGGATGATTATATAACGAAGCCCTTTCGCCCACTGAGTTAATTGCTCGGGTA

LysAlaGlnLeuArgArgTyrLysLysPheSerGlyValLysGluGlnAsnGluAsnVal
 ArgProSerCysAlaAspThrLysAsnSerValGlu***ArgSerArgThrLysMetLeu
 GlyProValAlaProIleGlnLysIleGlnTrpSerLysGlyAlaGluArgLysCysTyr
 AAGGCCAGTTGCGCCGATACAAAAAATTCAGTGGAGTAAAGGAGCAGAACGAAAATGTT

1900

IleValHisSerGlyLeuValIleAsnValAsnThrHisGluCysTyrLeuAsnGluLys
 SerSerThrProAlaLeuSerLeuMetLeuThrProMetSerValIle***ThrArgSer
 ArgProLeuArgProCysHis***Cys***HisPro***ValLeuSerGluArgGluAla
 ATCGTCCACTCCGGCCTTGTCATTAATGTAAACACCCATGAGTGTATCTGAACGAGAAG

GlnLeuSerLeuThrProThrGluPheSerIleLeuArgIleLeuCysGluAsnLysGly
 SerTyrProLeuLeuProProSerPheGlnTyrCysGluSerSerValLysThrArgGly
 ValIleProTyrSerHisArgValPheAsnThrAlaAsnProLeu***LysGlnGlyGlu
 CAGTTATCCCTTACTCCCACCGAGTTTTCAATACTGCGAATCCTCTGTGAAAACAAGGGG

2000

AsnValValSerSerGluLeuLeuPheHisGluIleTrpGlyAspGluTyrPheSerLys
 MetTrpLeuAlaProSerCysTyrPheMetArgTyrGlyAlaThrAsnIleSerAlaArg
 CysGly***LeuArgAlaAlaIleSer***AspMetGlyArgArgIlePheGlnGlnGlu
 AATGTGGTTAGCTCCGAGCTGCTATTTTCATGAGATATGGGGCCACGAATATTTTCAGCAAG

2100

SerAsnAsnThrIleThrValHisIleArgHisLeuArgGluLysMetAsnAspThrIle
 AlaThrThrProSerProCysIleSerGlyIleCysAlaLysLys***ThrThrProLeu
 GlnGlnHisHisHisArgAlaTyrProAlaPheAlaArgLysAsnGluArgHisHis***
 AGCAACAACACCATCACCGTGCATATCCGGCATTTCGCGCAAAAAATGAACGACACCATT

 AspAsnProLysTyrIleLysThrValTrpGlyValGlyTyrLysIleGluLys***Lys
 IleIleArgAsnIle***LysArgTyrGlyGlyLeuValIleLysLeuLysAsnLysLys
 SerGluIleTyrLysAsnGlyMetGlyGlyTrpLeuAsn***LysIleLysLys
 GATAATCCGAAATATATAAAAACGGTATGGGGGGTTGGTTATAAAATTGAAAAATAAAAA
 2200 . . .
 LysArgLeuPheGlnThrArgThrLysThrLeuHisValTyrArgCysAsnCysCysGly
 AsnAspTyrSerLysLeuGluArgLysLeuTyrMetTyrIleValAlaIleValValVal
 ThrThrIleProAsn***AsnGluAsnPheThrCysIleSerLeuGlnLeuLeuTrp***
 AAACGACTATTCCAACTAGAACGAAAACCTTTACATGTATATCGTTGCAATTGTTGTGGT

 SerAsnCysIleArgValValTyrSerPheAsnAspProArgGluThrTrpGlyLeuAsp
 AlaIleValPheValLeuTyrIleArgSerMetIleArgGlyLysLeuGlyAspTrpIle
 GlnLeuTyrSerCysCysIlePheValGln***SerGluGlyAsnLeuGlyIleGlySer
 AGCAATTGTATTCTGTGTGTATATTCGTTCAATGATCCGAGGGAACTTGGGGATTGGAT
 2300 . . .
 LeuLysTyrPheGlyLysGlnIle***LeuLysSerProGlyArgAspGluIleIleSer
 LeuSerIleLeuGluAsnLysTyrAspLeuAsnHisLeuAspAlaMetLysLeuTyrGln
 ValPheTrpLysThrAsnMetThrIleThrTrpThrArg***AsnTyrIleAsn
 CTTAAGTATTTTGGAAAACAAATATGACTTAAATCACCTGGACGCGATGAAATTATATCA
 2400

IlePheHisThrGluGlnTyrArgTyrLeuTyrLeuCysGlyAspCysHis***TyrSer
 TyrSerIleArgAsnAsnIleAspIlePheIleTyrValAlaIleValIleSerIleLeu
 IleProTyrGlyThrIle***IleSerLeuPheMetTrpArgLeuSerLeuValPheLeu
 ATATTCCATACGGAACAATATAGATATCTTTATTTATGTGGCGATTGTCATTAGTATTCT

TyrSerMetSerArgHisAlaPheLysIleArgLysIleLeu***ArgAspLysTyrArg
 IleLeuCysArgValMetLeuSerLysPheAlaLysTyrPheAspGluIleAsnThrGly
 PheTyrValAlaSerCysPheGlnAsnSerGlnAsnThrLeuThrArg***IleProAla
 TATTCTATGTGCGTCATGCTTTCAAATTCGCAAATACTTTGACGAGATAAATACCGG

2500

His***CysThrTyrSerGluArgArg***ThrAsn***AlaPheCysGlyAsnGlyCys
 IleAspValLeuIleGlnAsnGluAspLysGlnIleGluLeuSerAlaGluMetAspVal
 LeuMetTyrLeuPheArgThrLysIleAsnLysLeuSerPheLeuArgLysTrpMetLeu
 CATTGATGTACTTATTCAGAACGAAGATAAACAATTGAGCTTTCTGCGGAAATGGATGT

TyrGlyThrLysAlaGlnHisIleLysThrAspSerGlyLysAlaArgAlaGlyCysLys
 MetGluGlnLysLeuAsnThrLeuLysArgThrLeuGluLysArgGluGlnAspAlaLys
 TrpAsnLysSerSerThrHis***AsnGlyLeuTrpLysSerGluSerArgMetGlnSer
 TATGGAACAAAAGCTCAACACATTAAACGGACTCTGGAAAAGCGAGAGCAGGATGCAA

2600

AlaGlyArgThrLysLysLys***ArgCysTyrValLeuGlyAlaArgTyr***AsnAla
 LeuAlaGluGlnArgLysAsnAspValValMetTyrLeuAlaHisAspIleLysThrPro
 TrpProAsnLysGluLysMetThrLeuLeuCysThrTrpArgThrIleLeuLysArgPro
 GCTGGCCGAACAAAGAAAAAATGACGTTGTTATGTACTTGGCGCACGATATTAAACGCC

2700

ProTyrIleHisTyrArgLeuPheGluProAla***ArgGlySerArgHisAlaGlyArg
 LeuThrSerIleIleGlyTyrLeuSerLeuLeuAspGluAlaProAspMetProValAsp
 LeuHisProLeuSerValIle***AlaCysLeuThrArgLeuGlnThrCysArg***Ile
 CCTTACATCCATTATCGGTTATTTGAGCCTGCTTGACGAGGCTCCAGACATGCCGGTAGA

SerLysGlyLysValCysAlaTyrHisValGlyGlnSerValSerThrArgThrAlaAsn
 GlnLysAlaLysTyrValHisIleThrLeuAspLysAlaTyrArgLeuGluGlnLeuIle
 LysArgGlnSerMetCysIleSerArgTrpThrLysArgIleAspSerAsnSer***Ser
 TCAAAGGCAAAGTATGTGCATATCACGTTGGACAAAGCGTATCGACTCGAACAGCTAAT

2800

ArgArgValPhe***AspTyrThrVal***ProThrAsnAspAsnAlaAsnLysAsnAla
 AspGluPhePheGluIleThrArgTyrAsnLeuGlnThrIleThrLeuThrLysThrHis
 ThrSerPheLeuArgLeuHisGlyIleThrTyrLysArg***Arg***GlnLysArgThr
 CGACGAGTTTTTTGAGATTACACGGTATAACCTACAAACGATAACGCTAACAAAAACGCA

HisArgProIleLeuTyrAlaGlyAlaAspAspArg***IleLeuSerSerAlaPheArg
 IleAspLeuTyrTyrMetLeuValGlnMetThrAspGluPheTyrProGlnLeuSerAla
 ThrTyrThrIleCysTrpCysArgProMetAsnPheIleLeuSerPheProHis
 CATAGACCTATACTATATGCTGGTGCAGATGACCGATGAATTTTATCCTCAGCTTTCCGC

2900

ThrTrpLysThrGlyGlyTyrSerArgProArgGlySerAspArgValArgArgPro***
 HisGlyLysGlnAlaValIleHisAlaProGluAspLeuThrValSerGlyAspProAsp
 MetGluAsnArgArgLeuPheThrProProArgIle***ProCysProAlaThrLeuIle
 ACATGGAAAACAGGCGGTTATTCACGCCCCGAGGATCTGACCGTGTCCGGCGACCCTGA

3000

ThrArgGluSerLeuGlnHisPheGluLysArgArgCysIleGln***Gly***
 LysLeuAlaArgValPheAsnAsnIleLeuLysAsnAlaAlaAlaTyrSerGluAspAsn
 AsnSerArgGluSerLeuThrThrPhe***LysThrProLeuHisThrValArgIleThr
 TAAACTCGCGAGAGTCTTTAACAACATTTTGAAAAACGCCGCTGCATACAGTGAGGATAA

 GlnHisHis***HisTyrArgGlyProLeuArgGlyCysGlyValAsnArgIleGlnGlu
 SerIleIleAspIleThrAlaGlyLeuSerGlyAspValValSerIleGluPheLysAsn
 AlaSerLeuThrLeuProArgAlaSerProGlyMetTrpCysGlnSerAsnSerArgThr
 CAGCATCATTGACATTACCGCGGGCCTCTCCGGGGATGTGGTGTCAATCGAATTCAAGAA
 3100 . . .
 HisTrpLysHisProLysArg***AlaSerCysHisIle***LysValLeu***AlaGly
 ThrGlySerIleProLysAspLysLeuAlaAlaIlePheGluLysPheTyrArgLeuAsp
 LeuGluAlaSerGlnLysIleSer***LeuProTyrLeuLysSerSerIleGlyTrpThr
 CACTGGAAGCATCCCAAAGATAAGCTAGCTGCCATATTTGAAAAGTTCTATAGGCTGGA

 GlnPheSerPhePheArgTyrGlyTrpArgGlyThrTrpIleGlyAspCysLysArgAsn
 AsnSerArgSerSerAspThrGlyGlyAlaGlyLeuGlyLeuAlaIleAlaLysGluIle
 IleLeuValLeuProIleArgValAlaArgAspLeuAspTrpArgLeuGlnLysLysLeu
 CAATTCTCGTTCTTCCGATACGGGTGGCGCGGGACTTGGATTGGCGATTGCAAAAGAAAT
 3200 . . .
 TyrCysSerAlaTrpArgAlaAspLeuArgGlyLysLeu*****LeuTyrAspVal***
 IleValGlnHisGlyGlyGlnIleTyrAlaGluSerTyrAspAsnTyrThrThrPheArg
 LeuPheSerMetGluGlyArgPheThrArgLysAlaMetIleThrIleArgArgLeuGly
 TATTGTTTCAGCATGGAGGGCAGATTTACGCGGAAAGCTATGATAACTATACGACGTTTAG
 3300 . . .

GlyArgAlaSerSerAspAlaArgLeuGly*****LysGluValLeuArgAspValTyr
 ValGluLeuProAlaMetProAspLeuValAspLysArgArgSer***GluMetTyrIle
 SerPheGlnArgCysGlnThrTrpLeuIleLysGlyGlyProLysArgCysIle
 GGTAGAGCTTCCAGCGATGCCAGACTTGGTTGATAAAAGGAGGTCCTAAGAGATGTATAT

 AsnPheLeuGlyLysSerGlnGlyTyrLeuTyrPhePheLeuGlyAsn***GlnPheAsn
 IlePhe***GluAsnLeuLysValIlePheThrPheSer***GluIleAsnAsnLeuIle
 PhePheArgLysIleSerArgLeuSerLeuLeuPheLeuArgLysLeuThrIle***Tyr
 AATTTTTTAGGAAAATCTCAAGGTTATCTTTACTTTTTCTTAGGAAATTAACAATTTAAT
 3400
 IleLysLysArgLeuValLeuThrArg***Thr***TyrArgLysAsnGluProPheSer
 LeuArgAsnGlySerPheLeuHisGlyArgLeuAsnThrValArgThrSerArgPheArg
 GluThrAlaArgSerTyrThrValAspLeuIleProGluArgAlaValPheVal
 ATTAAGAAACGGCTCGTTCTTACACGGTAGACTTAATACCGTAAGAACGAGCCGTTTTTCG

 PhePheArgGluArgPheAspLysIleThrIleGlyIleProValLeuPheGlyAlaPhe
 SerSerGluLysAspLeuThrArgLeuProLeuAlaSerProPheTyrLeuValProphe
 LeuGlnArgLysIle***GlnAspTyrHisTrpHisProArgPheIleTrpCysLeuSer
 TTCTTCAGAGAAAGATTTGACAAGATTACCATTGGCATCCCCGTTTTATTGGTGCCTTT
 3500
 HisArgLysGlyTrpSer***Leu***IleThrSerAlaLeuLeuPheMetAspValSer
 ThrGluArgValGlyLeuAsnTyrGlu***HisArgHisTyrCysLeuTrpMet***Ala
 GlnLysGlyLeuValLeuIleMetAsnAsnIleGlyIleThrValTyrGlyCysGluGln
 CACAGAAAGGGTTGGTCTTAATTATGAATAACATCGGCATTACTGTTTATGGATGTGAGC
 3600

ArgMetArgGlnMetHisSerMetLeuPheArgLeuAlaLeuAlaLeuTrpGlnArg***
 Gly***GlyArgCysIleProCysSerPheAlaSerLeuTrpArgTyrGlyAsnAspAsn
 AspGluAlaAspAlaPheHisAlaLeuSerProArgPheGlyValMetAlaThrIleIle
 AGGATGAGGCAGATGCATTCCATGCTCTTTTCGCCTCGCTTTGGCGTTATGGCAACGATAA

LeuThrProThrCysArgAsnProThrProAsnProArgLeuSerIleAsnValSerVal
 ***ArgGlnArgValGlyIleGlnArgGlnIleArgAlaPheGlnSerMetTyrGlnCys
 AsnAlaAsnValSerGluSerAsnAlaLysSerAlaProPheAsnGlnCysIleSerVal
 TTAACGCCAACGTGTCGGAATCCAACGCCAAATCCGCGCCTTTCAATCAATGTATCAGTG

3700

TrpAspIleAsnGlnArgPheProProLeuPhePheLeuArg***ArgGluProVal***
 GlyThr***IleArgAspPheArgLeuTyrSerSerCysAlaGluGluSerArgCysGlu
 GlyHisLysSerGluIleSerAlaSerIleLeuLeuAlaLeuLysArgAlaGlyValLys
 TGGGACATAAATCAGAGATTTCCGCCTCTATTCTTCTTGCGCTGAAGAGAGCCGGTGTGA

AsnIlePheLeuProGluAlaSerAlaAlaIleIle***IleGlnLeuLeuLeuArgGlu
 IleTyrPheTyrProLysHisArgLeuGlnSerTyrArgTyrAsnCysCys***GluAsn
 TyrIleSerThrArgSerIleGlyCysAsnHisIleAspThrThrAlaAlaLysArgMet
 AATATATTTCTACCCGAAGCATCGGCTGCAATCATATAGATACAACTGCTGCTAAGAGAA

3800

TrpAlaSerLeuSerThrMetTrpArgThrArgArgIleAlaLeuProIleIleLeu***
 GlyHisHisCysArgGlnCysGlyValLeuAlaGly***ArgCysArgLeuTyrTyrAsp
 GlyIleThrValAspAsnValAlaTyrSerProAspSerValAlaAspTyrThrMetMet
 TGGGCATCACTGTCGACAATGTGGCGTACTCGCCGGATAGCGTTGCCGATTATACTATGA

3900

Cys***PheLeuTrpGlnTyrAlaThr***AsnArgLeuCysAlaLeuTrpLysAsnMet
 AlaAsnSerTyrGlySerThrGlnArgLysIleAspCysAlaLeuCysGlyLysThr***
 LeuIleLeuMetAlaValArgAsnValLysSerIleValArgSerValGluLysHisAsp
 TGCTAATTCTTATGGCAGTACGCAACGTAAAATCGATTGTGCGCTCTGTGGAAAAACATG

 IleSerGlyTrpThrAlaThrValAlaArgTyrSerAlaThr***GlnLeuValTrpTrp
 PheGlnValGlyGlnArgProTrpGlnGlyThrGlnArgHisAspSerTrpCysGlyGly
 PheArgLeuAspSerAspArgGlyLysValLeuSerAspMetThrValGlyValValGly
 ATTCAGGTTGGACAGCGACCGTGGCAAGGTACTCAGCGACATGACAGTTGGTGTGGTGG
 4000
 GluArgAlaArg***AlaLysArgLeuLeuSerGlyCysGluAspLeuAspValLysCys
 AsnGlyProAspArgGlnSerGlyTyr***AlaAlaAlaArgIleTrpMet***SerVal
 ThrGlyGlnIleGlyLysAlaValIleGluArgLeuArgGlyPheGlyCysLysValLeu
 GAACGGGCCAGATAGGCAAAGCGGTTATTGAGCGGCTGCGAGGATTTGGATGTAAAGTGT

 TrpLeuIleValAlaAlaGluVal***Arg***ThrMetTyrArgLeuMetSerCysCys
 GlyLeu***SerGlnProLysTyrArgGlyLysLeuCysThrVal*****ValAlaAla
 AlaTyrSerArgSerArgSerIleGluValAsnTyrValProPheAspGluLeuLeuGln
 TGGCTTATAGTCGCAGCCGAAGTATAGAGGTAACTATGTACCGTTTGATGAGTTGCTGC
 4100
 LysIleAlaIleSerLeuArgPheMetCysArgSerIleArgIleArgThrIleLeuSer
 Lys***ArgTyrArgTyrAlaSerCysAlaAlaGlnTyrGlyTyrAlaLeuTyrTyrGln
 AsnSerAspIleValThrLeuHisValProLeuAsnThrAspThrHisTyrIleIleSer
 AAAATAGCGATATCGTTACGCTTCATGTGCCGCTCAATACGGATACGCACTATATTATCA
 4200

AlaThrAsnLysTyrArgGlu***SerLysGluHisPheLeuSerIleLeuGlyAlaVal
 ProArgThrAsnThrGluAsnGluAlaArgSerIleSerTyrGlnTyrTrpAlaArgSer
 HisGluGlnIleGlnArgMetLysGlnGlyAlaPheLeuIleAsnThrGlyArgGlyPro
 GCCACGAACAAATACAGAGAATGAAGCAAGGAGCATTCTTATCAATACTGGGCGCGGTC

HisLeu***IleProMetSerTrpLeuLysHis***LysThrGlyAsnTrpAlaValPro
 ThrCysArgTyrLeu***ValGly***SerIleArgLysArgGluThrGlyArgCysArg
 LeuValAspThrTyrGluLeuValLysAlaLeuGluAsnGlyLysLeuGlyGlyAlaAla
 CACTTGTAGATACCTATGAGTTGGTTAAAGCATTAGAAAACGGGAACTGGGCGGTGCCG

4300

HisTrpMetTyrTrpLysGluArgLysSerPheSerThrLeuIleAlaProLysAsnGln
 IleGlyCysIleGlyArgArgGlyArgValPheLeuLeu***LeuHisProLysThrAsn
 LeuAspValLeuGluGlyGluGluGluPhePheTyrSerAspCysThrGlnLysProIle
 CATTGGATGTATTGGAAGGAGAGGAAGAGTTTTTCTACTCTGATTGCACCCAAAAACCAA

LeuIleIleAsnPheTyrLeuAsnPheLysGluCysLeuThr*****SerHisArgIle
 *****SerIlePheThr***ThrSerLysAsnAla***ArgAspAsnHisThrAlaTyr
 AspAsnGlnPheLeuLeuLysLeuGlnArgMetProAsnValIleIleThrProHisThr
 TTGATAATCAATTTTTTACTTAAACTTCAAAGAATGCCTAACGTGATAATCACACCGCATA

4400

ArgProIleIleProSerLysArgCysValIleProLeuLysLysProLeuLysThrVal
 GlyLeuLeuTyrArgAlaSerValAla***TyrArg***LysAsnHis***LysLeuPhe
 AlaTyrTyrThrGluGlnAlaLeuArgAspThrValGluLysThrIleLysAsnCysLeu
 CGGCCTATTATACCGAGCAAGCGTTGCGTGATACCGTTGAAAAAACCATTAAAACTGTT

4500

TrpIleLeuLysGlyAspArgSerMetAsnArgIleLysValAlaIleLeuPheGlyGly
 GlyPhe***LysGluThrGlyAla***IleGlu***LysLeuGlnTyrCysLeuGlyVal
 AspPheGluArgArgGlnGluHisGlu***AsnLysSerCysAsnThrValTrpGlyLeu
 TGGATTTTGAAAGGAGACAGGAGCATGAATAGAATAAAAGTTGCAATACTGTTTGGGGGT

CysSerGluGluHisAspValSerValLysSerAlaIleGluIleAlaAlaAsnIleAsn
 AlaGlnArgSerMetThrTyrArg***AsnLeuGln***Arg***ProLeuThrLeuIle
 LeuArgGlyAla***ArgIleGlyLysIleCysAsnArgAspSerArg***His*****
 TGCTCAGAGGAGCATGACGTATCGGTAAAATCTGCAATAGAGATAGCCGCTAACATTAAT

4600

LysGluLysTyrGluProLeuTyrIleGlyIleThrLysSerGlyValTrpLysMetCys
 LysLysAsnThrSerArgTyrThrLeuGluLeuArgAsnLeuValTyrGlyLysCysAla
 ArgLysIleArgAlaValIleHisTrpAsnTyrGluIleTrpCysMetGluAsnValArg
 AAAGAAAATACGAGCCGTTATACATTGGAATTACGAAATCTGGTGTATGGAAAATGTGC

GluLysProCysAlaGluTrpGluAsnAspAsnCysTyrSerAlaValLeuSerProAsp
 LysAsnLeuAlaArgAsnGlyLysThrThrIleAlaIleGlnLeuTyrSerArgArgIle
 LysThrLeuArgGlyMetGlyLysArgGlnLeuLeuPheSerCysThrLeuAlaGly***
 GAAAAACCTTGCGCGGAATGGGAAAACGACAATTGCTATTCAGCTGTACTCTCGCCGGAT

4700

LysLysMetHisGlyLeuLeuValLysLysAsnHisGluTyrGluIleAsnHisValAsp
 LysLysCysThrAspTyrLeuLeuLysArgThrMetAsnMetLysSerThrMetLeuMet
 LysAsnAlaArgIleThrCys***LysGluPro***Ile***AsnGlnProCys***Cys
 AAAAAAATGCACGGATTACTTGTTAAAAGAACCATGAATATGAAATCAACCATGTTGAT

4800

ValAlaPheSerAlaLeuHisGlyLysSerGlyGluAspGlySerIleGlnGlyLeuPhe
 ***HisPheGlnLeuCysMetAlaSerGlnValLysMetAspProTyrLysValCysLeu
 SerIlePheSerPheAlaTrpGlnValArg***ArgTrpIleHisThrArgSerVal***
 GTAGCATTTTCAGCTTTGCATGGCAAGTCAGGTGAAGATGGATCCATACAAGGTCTGTTT

GluLeuSerGlyIleProPheValGlyCysAspIleGlnSerSerAlaIleCysMetAsp
 AsnCysProValSerLeuLeu***AlaAlaIlePheLysAlaGlnGlnPheValTrpThr
 IleValArgTyrProPheCysArgLeuArgTyrSerLysLeuSerAsnLeuTyrGlyGln
 GAATTGTCCGGTATCCCTTTTGTAGGCTGCGATATTCAAAGCTCAGCAATTTGTATGGAC

4900

LysSerLeuThrTyrIleValAlaLysAsnAlaGlyIleAlaThrProAlaPheTrpVal
 AsnArg***HisThrSerLeuArgLysMetLeuGly***LeuLeuProProPheGlyLeu
 IleValAspIleHisArgCysGluLysCysTrpAspSerTyrSerArgLeuLeuGlyTyr
 AAATCGTTGACATACATCGTTGCGAAAAATGCTGGGATAGCTACTCCCGCCTTTTGGGTT

IleAsnLysAspAspArgProValAlaAlaThrPheThrTyrProValPheValLysPro
 LeuIleLysMetIleGlyArgTrpGlnLeuArgLeuProIleLeuPheLeuLeuSerArg
 *****Arg*****AlaGlyGlySerTyrValTyrLeuSerCysPheCys***AlaGly
 ATTAATAAAGATGATAGGCCGGTGGCAGCTACGTTTACCTATCCTGTTTTTGTTAAGCCG

5000

AlaArgSerGlySerSerPheGlyValLysLysValAsnSerAlaAspGluLeuAspTyr
 ArgValGlnAlaHisProSerVal***LysLysSerIleAlaArgThrAsnTrpThrThr
 AlaPheArgLeuIleLeuArgCysGluLysSerGln***ArgGlyArgIleGlyLeuArg
 GCGCGTTCAGGCTCATCCTTCGGTGTGAAAAAAGTCAATAGCGCGGACGAATTGGACTAC

5100

AlaIleGluSerAlaArgGlnTyrAspSerLysIleLeuIleGluGlnAlaValSerGly
 GlnLeuAsnArgGlnAspAsnMetThrAlaLysSer***LeuSerArgLeuPheArgAla
 Asn***IleGlyLysThrIle***GlnGlnAsnLeuAsn***AlaGlyCysPheGlyLeu
 GCAATTGAATCGGCAAGACAATATGACAGCAAAATCTTAATTGAGCAGGCTGTTTCGGGC

 CysGluValGlyCysAlaValLeuGlyAsnSerAlaAlaLeuValValGlyGluValAsp
 ValArgSerValValArgTyrTrpGluThrValProArg***LeuLeuAlaArgTrpThr
 ***GlyArgLeuCysGlyIleGlyLysGlnCysArgValSerCysTrpArgGlyGlyPro
 TGTGAGGTCGGTTGTGCGGTATTGGGAAACAGTGCCGCGTTAGTTGTTGGCGAGGTGGAC
 5200
 GlnIleArgLeuGlnTyrGlyIlePheArgIleHisGlnGluValGluProGluLysGly
 LysSerGlyCysSerThrGluSerPheValPheIleArgLysSerSerArgLysLysAla
 AsnGlnAlaAlaValArgAsnLeuSerTyrSerSerGlySerArgAlaGlyLysArgLeu
 CAAATCAGGCTGCAGTACGGAATCTTTCGTATTCATCAGGAAGTCGAGCCGGAAAAAGGC

 SerGluAsnAlaValIleThrValProAlaAspLeuSerAlaGluGluArgGlyArgIle
 LeuLysThrGlnLeu***ProPheProGlnThrPheGlnGlnArgSerGluAspGlyTyr
 ***LysArgSerTyrAsnArgSerArgArgProPheSerArgGlyAlaArgThrAspThr
 TCTGAAAACGCAGTTATAACCGTTCCCGCAGACCTTTCAGCAGAGGAGCGAGGACGGATA
 5300
 GlnGluThrAlaLysLysIleTyrLysAlaLeuGlyCysArgGlyLeuAlaArgValAsp
 ArgLysArgGlnLysLysTyrIleLysArgSerAlaValGluVal***ProValTrpIle
 GlyAsnGlyLysLysAsnIle***SerAlaArgLeu***ArgSerSerProCysGlyTyr
 CAGGAAACGGCAAAAAAATATATAAAGCGCTCGGCTGTAGAGGTCTAGCCCGTGTGGAT
 5400

MetPheLeuGlnAspAsnGlyArgIleValLeuAsnGluValAsnThrLeuProGlyPhe
 CysPheTyrLysIleThrAlaAlaLeuTyr***ThrLysSerIleLeuCysProValSer
 ValPheThrArg***ArgProHisCysThrGluArgSerGlnTyrSerAlaArgPheHis
 ATGTTTTTACAAGATAACGGCCGCATTGTACTGAACGAAGTCAATACTCTGCCCGGTTTC

 ThrSerTyrSerArgTyrProArgMetMetAlaAlaAlaGlyIleAlaLeuProGluLeu
 ArgHisThrValValIleProVal***TrpProLeuGlnValLeuHisPheProAsn***
 ValIleGlnSerLeuSerProTyrAspGlyArgCysArgTyrCysThrSerArgThrAsp
 ACGTCATACAGTCGTTATCCCCGTATGATGGCCGCTGCAGGTATTGCACTTCCCGAACTG
 5500
 IleAspArgLeuIleValLeuAlaLeuLysGly*****AlaTrpLys***AspLeuLeu
 LeuThrAla***SerTyr***Arg***ArgGlyAspLysHisGlyAsnArgIleTyrPhe
 ***ProLeuAspArgIleSerValLysGlyValIleSerMetGluIleGlyPheThrPhe
 ATTGACCGCTTGATCGTATTAGCGTTAAAGGGGTGATAAGCATGGAAATAGGATTTACTT

 Phe***MetLys***TyrThrValPheValGlyThrLeuAsnMetProLeuGlyIleIle
 PheArg***AsnSerThrArgCysSerLeuGlyArg***IleCysHisLeuGly***Phe
 LeuAspGluIleValHisGlyValArgTrpAspAlaLysTyrAlaThrTrpAspAsnPhe
 TTTTAGATGAAATAGTACACGGTGTTCGTTGGGACGCTAAATATGCCACTTGGGATAATT
 5600
 SerProGluAsnArgLeuThrValMetLys***IleAlaLeu***GlyHisThrSerTrp
 HisArgLysThrGly***ArgLeu***SerLysSerHisCysArgAspIleArgValGly
 ThrGlyLysProValAspGlyTyrGluValAsnArgIleValGlyThrTyrGluLeuAla
 TCACCGGAAAACCGGTTGACGGTTATGAAGTAAATCGCATTGTAGGGACATACGAGTTGG
 5700

LeuAsnArgPhe***ArgGlnLysAsnTrpLeuLeuProLysGlyThrAspCysPheTyr
 ***IleAlaPheGluGlyLysArgThrGlyCysTyrProArgValArgIleAlaSerMet
 GluSerLeuLeuLysAlaLysGluLeuAlaAlaThrGlnGlyTyrGlyLeuLeuLeuTrp
 CTGAATCGCTTTTGAAGGCAAAAGAACTGGCTGCTACCCAAGSGTACGGATTGCTTCTAT

GlyThrValThrValLeuSerValLeu***ThrValLeuCysAsnGlyLeuHisSerArg
 GlyArgLeuProSer***AlaCysCysLysLeuPheTyrAlaMetGlyCysThrAlaGly
 AspGlyTyrArgProLysArgAlaValAsnCysPheMetGlnTrpAlaAlaGlnProGlu
 GGGACGGTTACCGTCCTAAGCGTGCTGTAACTGTTTTATGCAATGGGCTGCACAGCCGG

5800

LysIleThr***GlnArgLysValIleIleProIleLeuThrGluLeuArg***PheGln
 Lys***ProAspLysGlyLysLeuLeuSerGlnTyr***ProAsn***AspAspPheLys
 AsnAsnLeuThrLysGluSerTyrTyrProAsnIleAspArgThrGluMetIleSerLys
 AAAATAACCTGACAAAGGAAAGTTATTATCCCAATATTGACCGAACTGAGATGATTTCAA

LysAspThrTrpLeuGlnAsnGlnAlaIleAlaAlaAlaValProLeuIleLeuArgPhe
 ArgIleArgGlyPheLysIleLysPro***ProArgGlnCysHis***SerTyrAlaLeu
 GlyTyrValAlaSerLysSerSerHisSerArgGlySerAlaIleAspLeuThrLeuTyr
 AAGGATACGTGGCTTCAAAATCAAGCCATAGCCGCGGCAGTGCCATTGATCTTACGCTTT

5900

IleAsp***ThrArgValSerLeuTyrGlnTrpGlyAlaAspLeuIleLeuTrpMetAsn
 SerIleArgHisGly***AlaCysThrAsnGlyGluProIle***PheTyrGly***Thr
 ArgLeuAspThrGlyGluLeuValProMetGlySerArgPheAspPheMetAspGluArg
 ATCGATTAGACACGGGTGAGCTTGTACCAATGGGGAGCCGATTTGATTTTATGGATGAAC

6000

AlaLeuIleMetArgGlnMetGluTyrHisAlaMetLysArgLysIleAlaAspValCys
 LeuSerSerCysGlyLysTrpAsnIleMetGln***SerAlaLysSerGlnThrPheAla
 SerHisHisAlaAlaAsnGlyIleSerCysAsnGluAlaGlnAsnArgArgArgLeuArg
 GCTCTCATCATGCGGCAAATGGAATATCATGCAATGAAGCGCAAATCGCAGACGTTTGC

AlaProSerTrpLysThrValGlyLeuLysHisIleAlaSerAsnGlyGlyThrMetTyr
 LeuHisHisGlyLysGlnTrpVal***SerIle***ProArgMetValAlaLeuCysIle
 SerIleMetGluAsnSerGlyPheGluAlaTyrSerLeuGluTrpTrpHisTyrValLeu
 GCTCCATCATGGAAAACAGTGGGTTTGAAGCATATAGCCTCGAATGGTGGCACTATGTAT

6100

***GluThrAsnHisThrProIleAlaIleLeuIleSerPrcLeuAsnLysLeuLeuThr
 LysArgArgThrIleProGln***LeuPhe***PheProArg***IleAsnPhe***Pro
 ArgAspGluProTyrProAsnSerTyrPheAspPheProValLys***ThrPheAsnArg
 TAAGAGACGAACCATACCCCAATAGCTATTTTGATTTCCTCCGTTAAATAAACTTTTAACC

ValAlaArgThrAsnTyrIleSer***LeuPheArgGlnGluThrArgArgMet***Leu
 LeuHisGlyGlnThrIle***AlaAsnSerPheGlyArgLysProAspValCysAsnTrp
 CysThrAspLysLeuTyrLysLeuThrLeuSerAlaGlyAsnProThrTyrValThrGly
 GTTGACGACGACAACTATATAAGCTAACTCTTTCGGCAGGAAACCCGACGTATGTAACG

6200

ValLeuArgGluPheIleTyrSerArg***Tyr***ArgCysLysAlaGluArgTyrCys
 PheLeuGlyAsnLeuTyrIleValAspSerIleGluAspValArgGlnSerAspIleAla
 Ser***GlyIleTyrIle*****IleValLeuLysMet***GlyArgAlaIleLeuArg
 GTTCTTAGGGAATTTATATATAGTAGATAGTATTGAAGATGTAAGGCAGAGCGATATTGC

6300

GlyHisTyrLeuArgAlaLeuArgGlnAspSerLeuIleIleArgLeuIleAla***Arg
 ValIleIleCysValArgCysGlyLysIleAla*****Asp***SerHisArgGly
 SerLeuSerAlaCysAlaAlaAlaArg***ProAspAsnLysThrAspArgIleGluGly
 GGTCATTATCTGCGTGCGCTGCGGCAAGATAGCCTGATAATAAGACTGATCGCATAGAGG

GlyGlyIleSerHisArgProLeuSerThrGlySerSerAlaSerLeuAsnSerAlaTrp
 ValValPheHisThrAlaHisCysGlnGlnAlaValGlnProArg***IleGlnHisGly
 TrpTyrPheThrProProIleValAsnArgGlnPheSerLeuValLysPheSerMetGly
 GGTGGTATTTACACCGCCCATTTGTCAACAGGCAGTTCAGCCTCGTTAAATTCAGCATGG

6400

ValSerLeuMetLysIleHisLeuHisTrp*****IleGln***GlyGluIle
 TyrHisLeu***LysPheIleTyrIleGlyAspAsnSerLysSerSerArgAlaLys***
 IleThrTyrGluAsnSerSerThrLeuValIleIleValAsnProValGlyArgAsnAsn
 GTATCACTTATGAAAATTCATCTACATTGGTGATAATAGTAAATCCAGTAGGGCGAAATA

IleAspCysAsnLeuArgGlyLysThrAlaGlnSerGlnThrArgLeuCysArgLeuArg
 LeuThrValIleTyrGlyAlaLysArgHisAsnLeuLysArgAspCysAlaVal***Gly
 LeuPheThrGlyGlnAsnGlyThrIleSerAsnGluIleValProPheLysGly
 ATTGACTGTAATTTACGGGGCAAACGGCACAAATCTCAAACGAGATTGTGCCGTTTAAGG

6500

GlyArgPhe***LysTyrPheIleLeuProThrIle***LeuArgArgArgLeuLysMet
 GluAspSerArgAsnIleSerTyrPheGlnLeuTyrSer***GlyGlyAsp***Lys***
 LysIleLeuGluIlePheHisThrSerAsnTyrIleValLysGluGluThrGluAsnGlu
 GGAAGATTCTAGAAATATTTACATACTCCAACATATAGTTAAGGAGGAGACTGAAAATG

6600

LysLysLeuPhePheLeuLeuLeuLeuLeuPheLeuIleTyrLeuGlyTyrAspTyrVal
 ArgSerCysPhePheTyrCysTyrCysTyrSer***TyrThr***ValMetThrThrLeu
 GluValValPhePheIleValIleValIleLeuAsnIleLeuArgLeu***LeuArg***
 AAGAAGTTGTTTTTTTTTATTGTTATTGTTATTCTTAATATACTTAGGTTATGACTACGTT

 AsnGluAlaLeuPheSerGlnGluLysValGluPheGlnAsnTyrAspGlnAsnProLys
 MetLysHisCysPheLeuArgLysLysSerAsnPheLysIleMetIleLysIleProLys
 SerThrValPheSerGlyLysSerArgIleSerLysLeuSerLysSerGlnArg
 AATGAAGCACTGTTTTCTCAGGAAAAAGTCGAATTTCAAATTATGATCAAAATCCCAA
 6700
 GluHisLeuGluAsnSerGlyThrSerGluAsnThrGlnGluLysThrIleThrGluGlu
 AsnIle***LysIleValGlyLeuLeuLysIleProLysArgLysGlnLeuGlnLysAsn
 ThrPheArgLys***TrpAspPhe***LysTyrProArgGluAsnAsnTyrArgArgThr
 GAACATTTAGAAAATAGTGGGACTTCTGAAAATACCCAAGAGAAAACAATTACAGAAGAA

 GlnValTyrGlnGlyAsnLeuLeuLeuIleAsnSerLysTyrProValArgGlnGluVal
 ArgPheIleLysGluIleCysTyr***SerIleValAsnIleLeuPheAlaLysLysCys
 GlyLeuSerArgLysSerAlaIleAsnGln*****IleSerCysSerProArgSerVal
 CAGGTTTATCAAGGAAATCTGCTATTAATCAATAGTAAATATCCTGTTCGCCAAGAAGTG
 6800
 SerGlnIleSerIleTyrLeuAsnMetThrAsn*****MetAspThrGlyCys
 GluValArgTyrArgGluPheIle***Thr***ArgIleAsnLysTrpIleArgValAla
 LysSerAspIleValAsnLeuSerLysHisAspGluLeuIleAsnGlyTyrGlyLeuLeu
 TGAAGTCAGATATCGTGAATTTATCTAAACATGACGAATTAATAAATGGATACGGGTTGC
 6900

LeuIleValIlePheIleCysGlnLysLys***HisLysAsnPheGlnArgTrpSerMet
 *****TyrLeuTyrValLysArgAsnSerThrLysIlePheArgAspGlyGln***
 AspSerAsnIleTyrMetSerLysGluIleAlaGlnLysPheSerGluMetValAsnAsp
 TTGATAGTAATATTTATATGTCAAAAGAAATAGCACAAAAATTTTCAGAGATGGTCAATG

 MetLeu***ArgValAlaLeuValIleLeuLeuLeuIleValAlaIleGluThrLeuMet
 CysCysLysGlyTrpArg***SerPheTyrTyr*****TrpLeuSerArgLeu*****
 AlaValLysGlyGlyValSerHisPheIleIleAsnSerGlyTyrArgAspPheAspGlu
 ATGCTGTAAAGGGTGGCGTTAGTCATTTTATTATTAATAGTGGCTATCGAGACTTTGATG
 7000 . . .
 SerLysValCysPheThrLysLysTrpGlyLeuSerMetProTyrGlnGlnValIleVal
 AlaLysCysAlaLeuProArgAsnGlyGly***ValCysLeuThrSerArgLeu*****
 GlnSerValLeuTyrGlnGluMetGlyAlaGluTyrAlaLeuProAlaGlyTyrSerGlu
 AGCAAAGTGTGCTTTACCAAGAAATGGGGGCTGAGTATGCCTTACCAGCAGGTTATAGTG

 SerIleIleGlnValTyrHis***Met***AspGlnAla***ArgLysTrpAsnGluPro
 Ala***PheArgPheIleThrArgCysArgIleLysLeuAspGluAsnGlyThrSerPro
 HisAsnSerGlyLeuSerLeuAspValGlySerSerLeuThrLysMetGluArgAlaPro
 AGCATAATTCAGGTTTATCACTAGATGTAGGATCAAGCTTGACGAAAATGGAACGAGCCC
 7100
 LeuLysGluSerGly***LysLysMetLeuGlyAsnThrGlySerPheTyrValIleGln
 ***ArgLysValAspArgArgLysCysLeuGluIleArgValHisPheThrLeuSerArg
 GluGlyLysTrpIleGluGluAsnAlaTrpLysTyrGlyPheIleLeuArgTyrProGlu
 CTGAAGGAAAGTGGATAGAAGAAAATGCTTGGAATACGGGTTTCATTTTACGTTATCCAG
 7200

ArgThrLysGlnSer***GlnGluPhe

GlyGlnAsnArgValAsnArgAsnSer

AspLysThrGluLeuThrGlyIleGln

AGGACAAAACAGAGTTAACAGGAATTC

. . 7227